A Survey of Electronic Drug Information Resources and Identification of Problems Associated with the Differing Vocabularies Used to Key Them

John Angelo Gnassi, MD, and G. Octo Barnett, MD
Massachusetts General Hospital Laboratory of Computer Science

Drug information resources are increasingly becoming electronically available. They differ in scope, granularity, and purpose. These considerations have shaped the selection of dissimilar drug name keys, complicating access. An abbreviated and simplified historical context of the development of official controlled vocabularies and their relationships is followed by a review of the kinds of information available in several electronic drug information resources. The key vocabularies used are discussed with examples. Problems using the differing terms of the resource vocabularies are identified.

Agreement on names for drugs seems a given at first glance. Physicians, nurses, and pharmacists use drug names many times daily in the practice of medicine. It's a legitimate question how they could all function together if they didn't use the same vocabulary. In the ideal world, we'd always clearly specify exactly which drug we mean with a term from an agreed-upon vocabulary, or with a long and sufficiently descriptive name to avoid ambiguity.

In the day to day practice of medicine we trade off specificity for brevity. Humans can handle ambiguity because usually we have contextual clues to help guide our interpretations. When reading a written prescription, we may already have the extrinsic information that a patient has a certain illness, or that the prescription was written in an outpatient clinic. We also can take advantage of local semantics in the prescription itself to narrow our list of possible drugs by the stated route of administration or dose size. For example, when a pharmacist in a hospital sees an order for **ACYCLOVIR** to be used intravenously, he knows to dispense **ACYCLOVIR SODIUM**, which is the salt used intravenously; **ACYCLOVIR** is a term for the chemical found in the tablets.

Because of advances in software technology, one might expect that the confusion about the communication of so basic a concept as the identity of a drug would be eliminated in the electronic sphere of medical information. But there is no all-encompassing master drug thesaurus containing a standardized controlled vocabulary for all potentially important levels of granularity.

Indeed there are several extant controlled drug vocabularies. This in itself creates problems because their domains overlap, they have varying scopes, and their terms describe drugs at different levels of specificity. Therefore there are not one-to-one and onto mappings from one vocabulary to another, complicating identification of the same concepts across vocabularies.

A controlled vocabulary is a defined set of terms, a dictionary, from which phrases are drawn to express concepts. *Controlled* means that there is some lexical uniformity enforced by only allowing choices from the predefined set of terms. As pharmaceuticals are continually developed, anything short of instantaneous update will lead to incompleteness. However, since it is essential both to choose relatively simple names and also to avoid ambiguity, choosing names for new entries is often difficult. Adding new terms and resolving conflicts or ambiguities is known as *maintaining* the controlled vocabulary. The reason for lack of an ideal universal system is in part related to the fact that maintaining a large vocabulary is a non-trivial task that requires continuing effort.

Part of the problem of inconsistency between vocabularies is the differing ideas of just what is meant by *drug*. A drug can be a mixture of related chemical substances, such as **DIGITALIS** or **CONJUGATED ESTROGENS**. Modern drugs are more commonly purified single chemical compounds, such as **DIGOXIN** or **ETHINYL ESTRADIOL**. Drugs can be considered conceptually as chemical, as above, or conceptually as formulation, as they exist in the pharmacy with inert ingredients such as stabilizers, binders, and coloring, as in **ACETAMINOPHEN TABLETS**. Some drugs are mixtures of unrelated chemical substances, such as **LIDOCAINE AND EPINEPHRINE INJECTION**. We also talk about groups of drugs explicitly, as

0195-4210/92/$5.00

when speaking of **PENICILLINS**, or even more broadly, **ANTIBIOTICS**. Sometimes these broad groupings are by chemical class, like **PHENOTHIAZINES**, sometimes by clinical action, like **ANTIPSYCHOTICS**, and sometimes by mechanism of action, such as **ANTIDOPAMINERGICS**. Sometimes we consider different salts of an active compound or its base to be clinically interchangeable. A prescription for **HYDROXYZINE** may be filled by a pharmacist with either **HYDROXYZINE HYDROCHLORIDE** or **HYDROXYZINE PAMOATE**. But we make a medical distinction with some, as between **NAPROXEN** and **NAPROXEN SODIUM**. Indeed, all of these meanings of *drug* are applicable to the terms used to key at least one of the vocabularies or resources discussed below.

OFFICIAL DRUG VOCABULARIES

Several large projects have attempted to officially formalize drug vocabularies on a national and international level. In 1953 the World Health Organization (WHO) began a global controlled drug vocabulary with the intent of facilitating the exchange of information on drugs worldwide. Its International Nonproprietary Name (INN) project goes on to this day [1]. The WHO determines an INN for a new drug and recommends this term to member states, but they are free to reject or accept each term individually. The INN vocabulary currently contains greater than 10,000 key terms at the chemical conceptual drug level.

Many countries have their own official vocabulary and associated organization to maintain it. A small sample includes the British Approved Name (BAN) vocabulary of the British Pharmacopoeia Commission (BPC), the vocabulary of the French Codex Commission (DCF; Dénomination Commune Française), and the Japanese Accepted Name (JAN) vocabulary of the Yakumu-Koho.

On the national level, the first controlled vocabulary in the United States (US) was the National Formulary [2] (NF). It was initially developed over a century ago and was actively used up until the early 1960's. It started as simply a catalog, without explicit and consistent rules for naming new drugs. It was incomplete and has been superseded. Some federal regulations still refer to the NF.

In 1961, the United States Adopted Names (USAN) program [1] was begun with the objective of producing and maintaining a complete, accurate, and convenient controlled vocabulary of medicinal chemicals, both prescription and over-the-counter (OTC), in use in the US. A USAN is a nonproprietary name, intended to be used without restriction as the identifier of a pharmacological substance (at the conceptual level) rather than as the identifier of specific formulations. USANs are particularly recommended for use in teaching, by hospital formularies, by drug companies (to uniquely identify the compound which they make, protecting their trademark; USANs are also required by federal law in all advertising and labels), and in journals and other professional communications. Initially USANs were adopted directly from the NF, but specific nomenclature guidelines and procedural rules have been developed and used for additions since then. The USAN doesn't span the set of all pharmaceuticals in the world, but does cover all those readily available in the US (as well as many that aren't commercially available, such as abandoned or investigational drugs).

Since 1962, the United States Food and Drug Administration (FDA) has given legal authority to the Department of Health and Human Services (HHS) to designate an official name for a drug. Since 1984, the FDA maintains that it will only use this route of neologism under unusual circumstances. Furthermore, it specifies that if it doesn't designate an official drug name, the USAN will be the official drug name. Thus, since the HHS designates exceedingly few drug names, the USAN is largely intended to be the official controlled drug vocabulary in the United States.

The USAN provides 7,781 names at the chemical conceptual level, but not further to the formulation level. For example, **ACETAZOLAMIDE** and **ACETAZOLAMIDE SODIUM, STERILE** are separate USAN terms (as well as distinct chemicals), but the granularity does not go on to the next (formulation) level of specifying that the product is, for example, a tablet, capsule, or intravenous solution.

When considering drugs just at the conceptual level, ideally we'd like a single unified vocabulary. To a more or less successful degree we already have identical conceptual drug names in official vocabularies throughout the world (although perhaps with spelling variants) [14]. In the US accord is promoted by standard procedural review in which proposed USANs are submitted for consideration to

the WHO, BPC, DCF, and FDA. Usually the final name chosen is the same. Most drug compendia draw their vocabularies from these sources, so the terms have a truly worldwide distribution. It is unfortunate that sometimes these organizations can't all agree, and different names are adopted for the same drug. For example, **ACETAMINOPHEN** (USAN) is **PARACETAMOL** (INN).

The vocabularies keyed by a chemical conceptual drug term (e.g. USAN: **CODEINE PHOSPHATE**) are appropriate for some applications where individual components are the items of interest. But some uses require a level of specificity at the combination-specific chemical conceptual level, e.g. to name a group of drugs containing only both acetaminophen and codeine, without reference to a physical form.

An example of this type is the FDA Standard Generic Name (SGN) vocabulary, used in the FDA "Orange Book" [5]. This is largely the USAN terms of all the active components of a medication, alphabetized, and separated by semicolons. The information contained is keyed by these terms, such as the single term **ACETAMINOPHEN; CODEINE PHOSPHATE.**

The USAN Council and the United States Pharmacopeial Convention (USPC) are closely related organizations. While the former is concerned only with nomenclature, the latter also provides guidelines for drug production, required degrees of purity, storage conditions, prescription, and more. Some USPC information is therefore appropriately keyed by a preparation-specific term, such as **ACETAMINOPHEN AND CODEINE PHOSPHATE TABLETS.** The USP vocabulary [3,4] is based nearly entirely on USAN terms, which are combined as needed with **AND**, comma, and usually one of 142 suffixes, such as **TOPICAL AEROSOL SOLUTION** or **SLOW-RELEASE CAPSULES.** Therefore the USP term uniquely identifies the physical forms or properties of product formulations for which the USPC has set standards.

RESOURCES AND VOCABULARIES

Much of the information mentioned above is available in electronic format. But there are other printed drug information resources and vocabularies now available electronically as well.

One of the USPC products, *USP Drug Information* [4] (USP DI), consists of tagged text fields, examples of which are "Chemical Name", "Usual Adult Dosage", "Drug Interactions", and "Side/Adverse Effects". There are a total of 156 tags which can be applied to either drug groups or individual preparations. The most finely grained information (pertaining to particular formulations) is keyed by the USP vocabulary. It is complete for practical day-to-day clinical prescribing purposes, encompassing 636 drug groups containing terms for 2076 unique preparations, but may lack abandoned or experimental drugs. The American Society of Hospital Pharmacists also publishes a similar resource, *AHFS Drug Information* [6].

For mapping between differing terms in some vocabularies, and to document which brand name preparations contain particular conceptual drugs, the *USAN and the USP Dictionary of Drug Names* [1] (UUDDN) is a valuable resource. The UUDDN contains, in tagged fields, all the USAN vocabulary and some of INN, BAN, and DCF vocabularies (where different), short common names, all former and current brand names in the US and Canada, Chemical Abstract Society (CAS) names and numbers, investigational drug codes, and others.

Physicians' GenRx [7] (formerly *Physicians' Generix*), is a compendium of prescription drug information (no OTC), published yearly since 1991. Most of the contained information is collected from US federal government databases and edited to remove redundancy and resolve inconsistencies, organized, and indexed. A partial list of contents includes average wholesale prices, therapeutic equivalents, the FDA approved prescribing information ("package insert"), all past and present US brand names, and some patent expiration dates. Each drug is tagged with categories, which consist of labels related to FDA approved indications, clinical effects, some clinical indications not approved by the FDA, pharmacology, sales volume, FDA approval date, and others. The information is keyed by the SGN vocabulary.

The *Physicians' Desk Reference* [8] contains individual FDA approved prescribing information sheets from each different manufacturer of a drug (if the manufacturer paid to be included). Some OTC drugs are included (still more are available in a smaller companion handbook). The information is usually keyed by brand name, but also sometimes by keys consisting of the combined names of closely related products from the same manufacturer (i.e. **CHILDREN'S TYLENOL ACETAMINOPHEN**

CHEWABLE TABLETS, ELIXIR, DROPS). Entries from different manufacturers of the same drug are highly (although curiously not completely) redundant.

The *Medical Letter's Drug Interaction Program* [9,10] (MLDIP) is a database of drug interactions, with references. Its information is keyed by a drug vocabulary apparently of their own design. Included are brand names, USANs, INNs, grouping terms such as **ERYTHROMYCINS**, and others.

COSTAR [11] is an ambulatory medical record system developed at Massachusetts General Hospital and in use for some 20 years. This system contains its own controlled drug vocabulary, which is maintained by the hospital. COSTAR key drug terms include common or trivial names (**ALBUTEROL INHALER**), USAN terms (**ASPIRIN**), brand names (**NICORETTE GUM**), and even broader terms (**CALCIUM SUPPLEMENT**, and **ASPIRIN WITH BARBITURATE**).

Using these resources together becomes problematic because of the disparate terminology used to key contained information. The following tables display a few examples of differing key terms for similar concepts in the COSTAR, INN, USAN, USP, SGN, MLDIP, and PDR vocabularies. From examining these and others, the following were noted:

- The INN term can simply be in basic disagreement with the other terms.
- Some vocabularies are more complete than others, preventing bi-directional onto mapping, which would hinder inverse lookups.
- The SGN term sometimes loses part of the name (the chemical salt portion) compared with the USAN terms for its components, which would impede simple lexical transformation between it and other corresponding terms.
- The inherently different levels of specificity in the vocabularies prevents one-to-one mapping.
- The PDR terms are often the most specific, but usually don't resemble the other terms.
- MLDIP and COSTAR terms don't share the same level of specificity, even internally.
- The differences in these medication names may appear to be trivial, but in the absence of medical knowledge and context it is difficult or very complicated to convert or map automatically between terms in these vocabularies.

To manage all of these levels of specificity and overlapping groupings requires both a large vocabulary and explicit chemical, pharmacological, and clinical attributes and relationships. The Semantic Net model of the UMLS™ project [12] could provide a starting framework. Alternatively, Evans, et al [13], have also discussed issues involved in mapping between medical vocabularies with similar discordance of scope and specificity, and suggest a hierarchical frame-based approach. Problems uncovered comparing these drug vocabularies argue for research into some unifying approach so that they may yield their true promise of making reference to drugs more clear, consistent, and concise. ▪

COSTAR:	ACETAMINOPHEN
INN:	PARACETAMOL
USAN:	ACETAMINOPHEN
USP:	ACETAMINOPHEN TABLETS
	ACETAMINOPHEN CAPSULES
	ACETAMINOPHEN ORAL SOLUTION
	ACETAMINOPHEN ORAL SUSPENSION
	ACETAMINOPHEN TABLETS (CHEWABLE)
	ACETAMINOPHEN WAFERS
	ACETAMINOPHEN SUPPOSITORIES
	ACETAMINOPHEN FOR EFFERVESCENT ORAL SOLUTION
SGN:	ACETAMINOPHEN
MLDIP:	ACETAMINOPHEN
PDR:	TYLENOL ACETAMINOPHEN CHILDREN'S CHEWABLE TABLETS & ELIXIR
	TYLENOL ACETAMINOPHEN CHILDREN'S SUSPENSION LIQUID
	TYLENOL, EXTRA STRENGTH, ACETAMINOPHEN ADULT LIQUID PAIN RELIEVER
	TYLENOL, EXTRA STRENGTH, ACETAMINOPHEN GELCAPS, CAPLETS, TABLETS
	TYLENOL, INFANTS' DROPS AND INFANTS' SUSPENSION DROPS
	TYLENOL, JUNIOR STRENGTH, ACETAMINOPHEN COATED CAPLETS, GRAPE AND FRUIT CHEWABLE TABLETS
	TYLENOL, REGULAR STRENGTH, ACETAMINOPHEN TABLETS AND CAPLETS
	APAP-ELIXIR

COSTAR:	ACYCLOVIR
INN:	ACYCLOVIR
	ACYCLOVIR SODIUM
USAN:	ACYCLOVIR
	ACYCLOVIR SODIUM
USP:	ACYCLOVIR CAPSULES
	ACYCLOVIR OINTMENT
	ACYCLOVIR SODIUM, STERILE
SGN:	ACYCLOVIR
MLDIP:	ACYCLOVIR
PDR:	ZOVIRAX CAPSULES
	ZOVIRAX OINTMENT 5%
	ZOVIRAX SUSPENSION
	ZOVIRAX TABLETS
	ZOVIRAX STERILE POWDER

COSTAR:	CODEINE
INN:	CODEINE PHOSPHATE
	CODEINE SULFATE
USAN:	CODEINE PHOSPHATE
	CODEINE SULFATE
USP:	CODEINE PHOSPHATE INJECTION
	CODEINE PHOSPHATE ORAL SOLUTION
	CODEINE PHOSPHATE TABLETS
	CODEINE PHOSPHATE SOLUBLE TABLETS
	CODEINE SULFATE TABLETS
	CODEINE SULFATE SOLUBLE TABLETS
	CODEINE SULFATE SOLUBLE TABLETS
SGN:	CODEINE PHOSPHATE
MLDIP:	CODEINE
PDR:	CODEINE PHOSPHATE INJECTION (availability info only)
	CODEINE SULFATE TABLETS (availability info only)
	CODEINE PHOSPHATE IN TUBEX (availability info only)

COSTAR:	(not in vocabulary)
INN:	(no multidrug terms in vocabulary)
USAN:	(no multidrug terms in vocabulary)
USP:	BUTALBITAL, ASPIRIN, CODEINE PHOSPHATE, AND CAFFEINE CAPSULES
	BUTALBITAL, ASPIRIN, CODEINE PHOSPHATE, AND CAFFEINE TABLETS
SGN:	ASPIRIN; BUTALBITAL; CAFFEINE; CODEINE
MLDIP:	(not in vocabulary)
PDR:	FIORINAL WITH CODEINE CAPSULES

Acknowledgments

Dr. Gnassi is supported by an NLM Medical Informatics Training Fellowship (1-T15-LM07092-01). This work was supported in parts by grants from the NLM (UMLS N01-LM-1-3538) and the Hewlett Packard Corporation.

Reference

[1]. USAN and the USP Dictionary of Drug Names, Carolyn A. Fleeger, ed., United States Pharmacopeial Convention, 1992 (electronic and print).

[2]. The National Formulary, 15th ed., United States Pharmacopeial Convention, 1980.

[3]. The United States Pharmacopeia, 20th revision, United States Pharmacopeial Convention, 1980.

[4]. USP Drug Information, vols. I and II, United States Pharmacopeial Convention, 1990 (electronic and print).

[5]. Approved Drug Products with Therapeutic Equivalence Evaluations ("Orange Book"), Food and Drug Administration, 1992.

[6]. AHFS Drug Information, Gerald K. McEvoy, ed., The American Society of Hospital Pharmacists, 1993.

[7]. Physicians' GenRx, Data Pharmaceutica, Inc., Philip L. Denniston, Jr., ed., 1993 (electronic and print).

[8]. Physicians' Desk Reference, 47th ed., Medical Economics Data, 1993 (electronic and print).

[9]. D. Gardner, M. Rizack, Software Distribution and Electronic Publishing of Adverse Drug Interaction Information, *SCAMC Proceedings*, 1987, p 726-727.

[10]. The Medical Letter's Drug Interaction Program, The Medical Letter, 1990 (software).

[11]. G.O. Barnett, The Application of Computer Based Medical Record Systems in Ambulatory Practice, *New England Journal of Medicine*, 1984;310:1643-1650.

[12]. UMLS Knowledge Sources - 3rd Experimental Edition, National Library of Medicine, August 1992.

[13]. D.A. Evans, D.J. Rothwell, I.A. Monarch, R.G. Lefferts, R.A. Cote, Toward Representations for Medical Concepts, *Medical Decision Making* [JC:ma8] 11(4 Suppl):S102-128, 1991 Oct-Dec.

[14]. Index Nominum, Swiss Pharmaceutical Society, eds., Medpharm, 1992.

[15]. Goodman and Gilman's The Pharmacological Basis of Therapeutics, 6th ed., Gilman, Goodman, Gilman, eds., Macmillan, 1980.